T. Tun*
Y. Kubagawa*
G. Dennis
P.D. Burrows
M.D. Cooper
H. Kubagawa

Genomic structure of mouse PIR-A6, an activating member of the paired immunoglobulin-like receptor gene family

Key words:
immunoglobulin-like receptors; initiator; innate immune receptors; short intron genes

Acknowledgments:
The authors thank Drs Fyodor A. Kondrashov and Elliot Lefkovitz for helpful discussion and Mr Dong-Won Kang and Mr Craig L. Maynard for technical assistance. Some of the data in this publication were generated through the use of Celera Discovery System and Celera's associated databases. This work was supported in part by National Institute of Allergy and Infectious Diseases Grants, AI 42127 (HK), AI 48098 (PDB) and AI 39816 (MDC).

Abstract: The gene for one of the activating members of the paired Ig-like receptor family, *Pira6*, was isolated from a genomic library and sequenced. The first of 9 exons in the ~8.2 kb *Pira6* gene encodes the 5′ untranslated region, the translation initiation site, and approximately half of the signal sequence. The second exon encodes the rest of the signal sequence, exons 3–8 each encode a single Ig-like extracellular domain, and exon 9 encodes the transmembrane region, cytoplasmic tail and 3′ UTR with four polyadenylation signals and six mRNA instability sequences. A soluble form of PIR-A6 may be generated by alternative splicing. The exonic sequences account for ~42% of the *Pira6* gene and ~34% for the single inhibitory *Pirb* gene, thus defining *Pira* and *Pirb* as genes with relatively short intronic sequences. Extensive sequence homology was found between *Pira6* and *Pirb* from ~2 kb upstream of the ATG initiation site to the beginning of intron 8. The *Pir* genes appear to be distributed in three regions of the proximal end of chromosome 7 based on the present data and an analysis of currently available mouse genomic sequence databases. One region contains a single *Pir* gene which is almost identical to *Pira6*, and the other two contain multiple *Pir* genes in opposite transcriptional orientations. Potential binding sites for hemopoiesis-specific and ubiquitous transcription factors were identified upstream of the *Pira6* transcription start sites that reside within the initiator consensus sequence motif. These results provide important clues to the coordinate regulation observed for PIR-A and PIR-B expression during hematopoiesis.

Authors' affiliations:

T. Tun[1],
Y. Kubagawa[1],
G. Dennis[2],
P.D. Burrows[2],
M.D. Cooper[1,4],
H. Kubagawa[3]

[1]Department of Medicine,

[2]Department of Microbiology,

[3]Department of Pathology, Division of Developmental and Clinical Immunology, University of Alabama at Birmingham,

[4]Howard Hughes Medical Institute, Birmingham, AL

Correspondence to:
H. Kubagawa
Division of Developmental Clinical Immunology
University of Alabama at Birmingham
WTI Rm 378
Birmingham
AL 35294–3300
USA
Tel: +205 975 7201
Fax: +205 975 7218
e-mail: hkubagawa@ccc.uab.edu

The paired immunoglobulin-like receptors of activating (PIR-A) and inhibitory (PIR-B) isoforms were originally identified in mice on the basis of limited homology with the human IgA Fc receptor (FcαR) (1, 2). PIR-A and PIR-B are cell surface glycoproteins containing similar extracellular regions with six Ig-like domains but distinctive transmembrane and cytoplasmic regions. The PIR-A protein has a short cytoplasmic tail and a charged arginine residue in the transmembrane domain that facilitates non-covalent association with an adaptor protein, the Fc

Received 19 September 2002, revised 24 December 2002, accepted for publication 28 December 2002

Tissue Antigens. ISSN 0001-2815

Tissue Antigens 2003 **61:** 220–230
Printed in Denmark. All rights reserved

*The first two authors contributed equally to this work

receptor common γ chain containing tyrosine-based activation motifs, to form a cell activation complex (3–6). In contrast, the PIR-B receptor has an uncharged transmembrane segment and a relatively long cytoplasmic tail containing three functional immunoreceptor tyrosine-based inhibitory motifs, thereby negatively regulating cellular activity via the SHP-1 and SHP-2 tyrosine phosphatases (7–12). PIR-A and PIR-B proteins are expressed by many hematopoietic cell types, including B lymphocytes, dendritic cells, monocyte/macrophages, granulocytes, mast cells, and megakaryocyte/platelets (4). The PIR ligands are still unknown, but the cellular distribution of the receptors suggests that PIR may function in both host defense and early hematopoiesis. Recent studies of PIR-B deficient mice revealed that PIR-B plays a critical *in vivo* regulatory role in B cell responses and dendritic cell maturation, thereby affecting the balance of Th1/Th2 immune responses (13).

The *Pir* gene family is located at the proximal end of chromosome 7 (1, 14, 15) in a region syntenic with the human chromosome 19q13 region, where a cluster of structurally related gene families called the leukocyte receptor complex (LRC) resides (16–20). The LCR genes include the platelet collagen receptor glycoprotein VI, the NK triggering receptor (NKp46), the FcαR, the killer Ig-like receptors (KIR), the leukocyte-associated Ig-like receptors (LAIR), and the closest PIR homolog the Ig-like transcripts (ILT) (also called leukocyte Ig-like receptors (LIR) and monocyte/macrophage Ig-like receptors). The genomic organization of the human *Ilt/Lir* genes and their LRC relatives has been well characterized, but much less is known about the organization of the mouse *Pir* gene region. In previous studies, PIR-B was found to be invariant and encoded by a single *Pirb* gene. On the other hand, PIR-A cDNA clones displayed sequence and domain structural variability in their extracellular regions and were predicted to be encoded by 6–10 *Pira* genes (1, 14, 15). However, genomic DNA blot analysis with a PIR-A specific probe revealed a relatively simple band pattern (less than four DNA fragments), possibly indicating conservation of restriction enzyme sites flanking the *Pira* genes (1, 14). The exon/intron organization and the promoter region of the single *Pirb* gene have been characterized (14, 15), but information about the *Pira* genes has been lacking. The sequence, exon/intron organization and transcription start sites of the *Pira6* gene have been determined in the present study to allow a comparison of the two types of *Pir* genes and their potential transcriptional regulatory motifs.

Materials and methods

Isolation of the PIR-A6 genomic clone

A P1 library (ES-129/OLA P1-4537, Genome System Inc., St. Louis, MO) containing genomic DNA inserts of 75–100 kb was constructed by partial digestion of 129/OLA mouse DNA with the *Sau*3AI restriction endonuclease prior to ligation into the *Bam*HI site of the pAd10SacII vector (21). Screening for P1 clones that contained *Pira* genes was performed by polymerase chain reaction (PCR) using a set of PIR-A specific primers: forward 5′-cccacagagaaccaggat-3′ and reverse 5′-cccagagtgtagaacattgaagatg-3′, which correspond to the short extracellular membrane proximal/transmembrane region and the 3′ untranslated region (UTR) of PIR-A1, respectively. PCR products with the expected size of ∼150 bp were amplified from four P1 clones, and one of them (P16335) was subjected to further characterization. To obtain a complete PIR-A genomic clone, the Expand high fidelity PCR System (Roche Diagnostics Corp., Indianapolis, IN) was employed using P16335 DNA as a template, a mixture of *Taq* and Pwo DNA polymerases, the latter of which has a 3′ to 5′ exonuclease proofreading activity, and the following primers: forward, 5′-agatgccatgtcctgcacctt-3′, corresponding to the 5′ UTR and signal sequence (SS) common to PIR-A and PIR-B, and reverse, 5′-aatggcagtaggctaatacatacacccgaccctcct-3′, corresponding to the 3′ UTR of PIR-A1 (1). The PCR was carried out in two stages: (i) initial 2 min denaturation at 92°C, followed by 10 cycles of 92°C for 10 s denaturation, 65°C for 30 s annealing and 68°C for 6 min extension, and (ii) 20 cycles of 92°C for 10 s denaturation, 65°C for 30 s annealing and 68°C for 20 min extension with a 20 section increase per cycle. The resultant PCR product of ∼7 kb was subcloned into the pCR-XL-TOPO vector (P16335-XL4).

Genomic sequencing

To sequence the P16335-XL4 insert DNA, PCR amplification was used to generate overlapping DNA fragments. Sense and antisense primers corresponding to the 5′-and 3′ end of each of the six extracellular Ig-like domains, the membrane proximal extracellular and the cytoplasmic regions were designed on the basis of the previously obtained sequence results from the *Pirb* and partial *Pira* genomic clones and the PIR cDNA clones. The resultant PCR products were subcloned into pCR2.1-TOPO-TA vectors and sequenced on both strands by the dideoxy chain-termination method using SequiTherm EXCEL II DNA polymerase (Epicentre Tech., Madison, WI) and an automated sequencer (Li-Cor, Lincoln, NE).

Isolation of 5′-and 3′ untranslated regions of *Pira6*

As the P16335-XL4 genomic clone did not contain the entire 5′-and 3′ *Pira* UTR, two *Hin*dIII fragments, ∼9 and ∼7.5 kb, containing, respectively, the 5′-and 3′ flanking region of the P16335-XL4 gene, were isolated from the original P16335 genomic clone and subcloned into pBluescript vectors (P16335–5′HIII and P16335–3′HIII) prior to

nucleotide sequence analysis. One kb of the 5′ flanking region of the ATG translation initiation site in the P16335–5′HIII was searched for putative transcription factor binding motifs by using the transcriptional factor site database TRANSFAC (22). The nucleotide sequences of these two genomic fragments were used to validate the sequence of the P16355-XL4.

Primer extension analysis

To determine the transcription start site of *Pir* genes, primer extension analysis was performed according the manufacturer's recommendation (Promega, Madison, WI), with some modifications. Briefly, PIR mRNA in 0.5–1 μg of poly(A)$^+$ mRNA isolated from adult BALB/cBy spleen and the WEHI3 macrophage cell line were extended toward to the 5′ end with the IRDye-700 labeled reverse primer 5′-aagacagagcagggctgtg-3′, which corresponds to the 3′ end of the SS in the first PIR-A exon, and avian myeloblastosis virus reverse transcriptase. The extended cDNA products were resolved by 7 M urea/3.7% polyacrylamide gel electrophoresis along with nucleotide sequencing products from the dideoxy-chain termination reaction of the P16335–5′HindIII DNA. Alternatively, 5′-rapid amplification of cDNA end (RACE) PCR was performed according to the manufacturer's recommendation (Clontech Laboratory, Palo Alto, CA) using the reverse primer 5′-caagacagagcagggctgtgaaggtgcagg-3′ corresponding to the same SS end.

Sequence analysis

Based on the previous sequence data of the PIR-A/PIR-B cDNA clones and the *Pirb* genomic clone and the present *Pira6* sequence data, repeat sequences in the PIR-A genomic sequence were identified using the CENSOR (http://www.girinst.org/Censor_Server.html) (23) and Repeat Masker (Smit, AFA & Green, P at http://ftp.genome.washington.edu/RM/RepeatMasker.html) websites. To characterize the *Pir* locus in the mouse genome, the genomic databases of Celera (http://www.celera.com) and the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html) were used to construct the map of the mouse *Pir* locus.

Results

Cloning strategy and identification of the *Pira6* gene

When a 129/OLA mouse P1 genomic library containing insert DNA with an average size of ~90 kb was screened by PCR using PIR-A specific primers, four PIR-A containing clones were identified. Restriction enzyme digestion of these clones followed by hybridization with the PIR-A specific probe revealed three different DNA mobility patterns. One of the clones (P16335) was subjected to long range PCR using a set of primers corresponding to the 5′UTR/SS and the 3′UTR of PIR-A. The resultant PCR product (P16335-XL4) of ~7 kb provided the PIR-A nucleotide sequence from the translation start site to the 3′UTR. The P16335-XL4 sequence was found to most closely resemble (99%) the BALB/c-derived PIR-A6 cDNA sequence, although the latter cDNA does not encode the fifth amino terminal Ig-like domain (1). Additional sequence for the complete 5′-and 3′ flanking regions was obtained from ~9 kb and ~7.5 kb HindIII DNA fragments of the original P16355 genomic clone (Fig. 1). In view of the possibility that the P16335 clone might contain multiple *Pira* genes, 23 subclones with an ~9 kb insert DNA and seven subclones with an ~7.5 kb insert DNA were sequenced. Surprisingly, all of these subclones were identical, thereby indicating that a single *Pira* gene, *Pira6*, is present in the P16335 P1 clone. The nucleotide sequence data from these two genomic fragments validated the P16335-XL4 sequence.

Organization of *Pira6* exons and introns

The *Pira6* gene (GenBank accession number AY216658), ~8.2 kb in size, contains nine exons (Fig. 1). The exonic sequences account for ~42% of the *Pira6* gene, a value slightly higher than the ~34% for the *Pirb* gene, thus defining *Pirb* and *Pirb* as genes with relatively short intronic sequences. The first *Pira6* exon, denoted as 5′UT/SS1, encodes the 5′UTR, the ATG translation initiation codon, and approximately half of the signal sequence. The second exon (SS2) is 36 bp long and encodes the remaining part of the signal sequence including the predicted signal cleavage site between threonine (Thr) and glycine (Gly). Exons 3–8 (EC1 to EC6) each encode a single Ig-like extracellular domain. Exon 9 (TM/CY/3′UT), the largest exon (1514 bp), encodes the short membrane proximal extracellular, transmembrane, cytoplasmic and 3′UT regions. In the 3′ UTR, there are four classical polyadenylation signals (AATAAA) and six mRNA instability motifs (ATTTA) (24). All of the *Pira6* exon/intron boundaries conform to the gt-ag splice-site consensus rule and are phase I, in that they occur after the first nucleotide of the triplet codon, a feature characteristic of members of the Ig gene superfamily (Table 1 and Fig. 2).

Possible soluble form of PIR-A6

Two types of PIR-A cDNA clones were identified in previous studies (1): one encoding a 28 amino acid (aa) membrane proximal extracellular region (ECmp) and the other an 11 aa ECmp. This difference

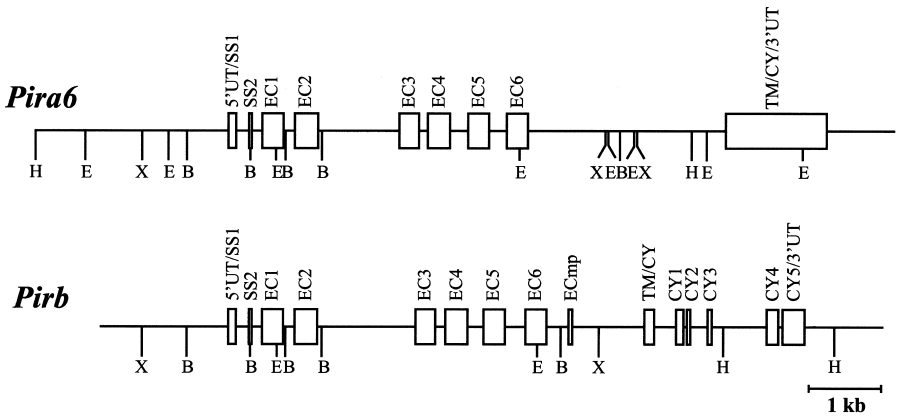

Fig. 1. Schematic presentation of *Pira6* and *Pirb* genes. The exon (boxes) organization of the *Pira6* and *Pirb* genes is drawn to the scale indicated, along with the restriction enzyme sites: B = *Bam*HI, E = *Eco*RI, H = *Hin*dIII and X = *Xba*I. Exons encoding particular regions of the receptors are denoted as follows: the 5′ untranslated (5′UT), the signal peptide (SS1 and SS2), the extracellular Ig-like domains (EC1 to EC6), the membrane proximal (ECmp), the transmembrane (TM), the cytoplasmic (CY), and the 3′ untranslated(3′UT) regions.

was predicted to result from alternative splicing of a mini-exon encoding a 17-aa ECmp. However, inspection of the 3′ flanking region of the EC6 exon of *Pira6* did not reveal a corresponding mini-exon; instead, we noted a large potential exon that would encode a soluble PIR-A6 form. This exon is located 233 bp downstream of the EC6 exon (Fig. 2). As it is flanked at its 5′ end with 'ag' as the 3′ splice acceptor signal, alternative splicing may occur between the EC6 and this exon. However, this putative exon would encode 39 aa instead of 17 aa, because the predicted splice donor site is 'at' rather than 'gt', hence extending the open reading frame an additional 22 amino acids before a stop codon. The resultant 39-aa region would be hydrophilic according to Kyte-Doolittle analysis (25) and would not contain a cleavage signal for glycosyl phosphatidylinositol linkage to the plasma membrane according to the GPI prediction website (http://mendel.imp.univie.ac.at/gpi/cgi-bin/gpi_pred.cgi) (26, 27). Five potential polyadenylation signals (aataaa) are present beyond the stop codon. Collectively, these features suggest that alternative splicing may generate a soluble form of PIR-A6.

As the predicted ECmp mini-exon was not identified in the P16335-XL4 genomic clone, we performed a BLAST search (28) of the Celera mouse genome database. Among seven genomic sequences identified on the basis of high homology (> 90%) to the EC6 and intron 9 regions of the *Pira6* gene, three appear to contain a functional ECmp mini-exon flanked with 'ag' and 'gt' at the 5′-and 3′ ends, respectively. Interestingly, the remainder had the same 'ag' and 'at' sequences at the corresponding sites as did the *Pira6* gene. Not all of these Celera sequences have been assigned yet to a specific chromosomal region, so this database was not helpful in computer-assisted identification of these *Pira* genes.

Exon-intron boundaries of the *Pira6* gene

Exon	Size (bp)	5′-Splice Donor*		Intron	Size (bp)	3′-Splice Acceptor**		Amino Acid
5′UT/SS1	99	CTCTGTCTT*G***	**gt**gagatgtg	1	154	tctcttcc**ag**	*GA*CTGACTCT	Gly
SS2	36	GTGCTGACA*G*	**gt**gagtgtgt	2	181	ttacttgc**ag**	*GG*TCCCTCCC	Gly
EC1	285	GTGGTGACA*G*	**gt**gagagaca	3	149	ttctctcc**ag**	*GA*CACTACTG	Gly
EC2	306	CTGGTCTCA*G*	**gt**gaggagcc	4	1106	ttctatgt**ag**	*GT*AATCTCCA	Gly
EC3	285	GTGGTGACA*G*	**gt**gaggggac	5	126	tctctcct**ag**	*GA*ATCTATGA	Gly
EC4	309	CTCATCTCA*G*	**gt**aagtcccc	6	228	ttccctct**ag**	*GG*CTGTCCAA	Gly
EC5	297	CTGATCACA*G*	**gt**gaggaact	7	235	ttctctttt**ag**	*GA*CAGCTCCC	Gly
EC6	303	ACAGTCTCA*G*	**gt**gaggtgat	8	2572	aataccac**ag**	*CC*ACAGAGAA	Ala
TM/CY/3′UT	1514	TTTAATACTA						

*Exon and intron sequences are shown in upper and lower case, and the splice donor (5′) and acceptor (3′) sequences in bold
**The exon boundary codons are in italics

Table 1

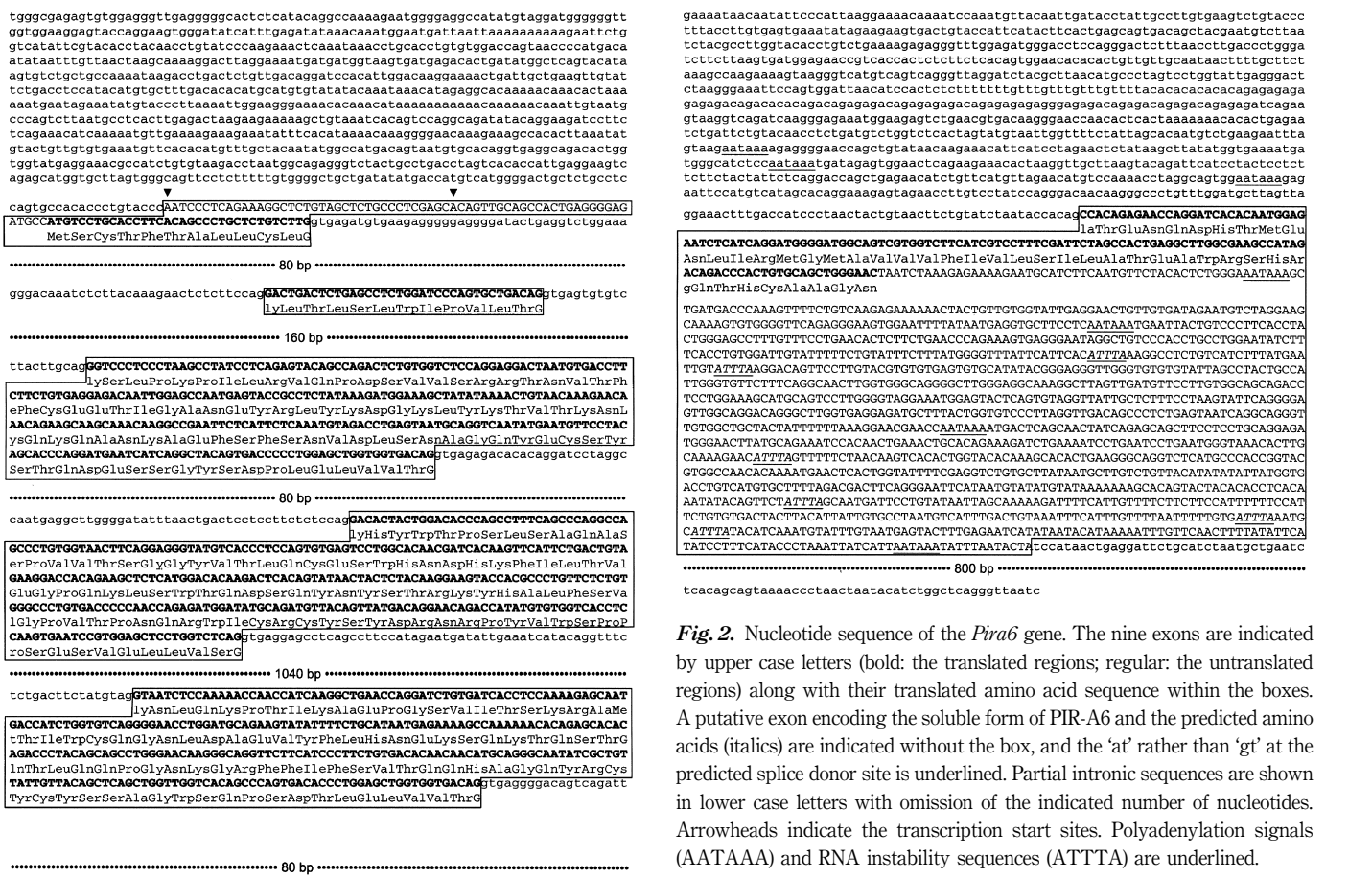

Fig. 2. Nucleotide sequence of the *Pira6* gene. The nine exons are indicated by upper case letters (bold: the translated regions; regular: the untranslated regions) along with their translated amino acid sequence within the boxes. A putative exon encoding the soluble form of PIR-A6 and the predicted amino acids (italics) are indicated without the box, and the 'at' rather than 'gt' at the predicted splice donor site is underlined. Partial intronic sequences are shown in lower case letters with omission of the indicated number of nucleotides. Arrowheads indicate the transcription start sites. Polyadenylation signals (AATAAA) and RNA instability sequences (ATTTA) are underlined.

Identification of *Pira6* transcription start sites

The largest PIR cDNA clone identified in our previous study was found to contain a 5′UTR of ~40 bp (1), thus suggesting that the transcription start site is in close proximity to the translation start site (1). To define the transcription start site of the *Pir* gene more precisely, we performed a primer extension analysis using mRNA and reverse transcriptase along with the 5′ flanking DNA fragment of *Pira6* as a control. Two different sized cDNA products were produced by the reverse primer, indicating *Pira6* transcription initiation 65 bp and 28 bp upstream of the translation start site (see Fig. 2). The same transcription initiation sites were identified by 5′ RACE analysis. The primary PIR-A6 transcript is thus ~8200 bp, which is slightly larger than the ~7900 bp primary PIR-B transcript. The *Pira6* gene, like *Pirb*, does not contain a classical TATA box upstream of the transcription initiation site. Inspection of the nucleotide sequence around the two start sites indicated that the most distal resides within an initiator consensus sequence $YYA_{+1}NWYY$, where Y, N and W represent C or T, any, and A or T nucleotide, respectively, and the subscript +1 indicates the position of the transcription start site ([29]; see Fig. 2). The proximal site is also

flanked by sequence similar to, but not perfectly matched with, the initiator consensus sequence. Analysis of the region upstream of the *Pira6* transcription start sites identified a number of sequence motifs that are predicted to bind various transcription factors (22). These include hemopoiesis-specific factors (e.g., MZF-1, c-Myb, Ikaros 2, NF-Y, GATA, C/EBPβ (30–32)); and ubiquitous factors (e.g., AP-1 (Jun/Fos), NF-κB) (Table 2).

Pira6 and *Pirb* comparison

When the genomic sequence of *Pira6* was compared with that of *Pirb* by dot matrix analysis, extensive sequence homology was evident from the 5′ ends, corresponding to the ∼2 kb 5′ flanking region, through the beginning of intron 8 downstream of the EC6 exon (Fig. 3). Notably, there was poor alignment around the ∼4200 bp position of *Pira6*, a region corresponding to intron 4 between EC2 and EC3. This finding is in agreement with the presence of ∼210 bp of additional sequence in the *Pirb* intron 4 (see Table 3). The numerical comparison of each exon and intron also indicates a high sequence homology (> 80% similarity) between *Pira6* and *Pirb* in the 5′ flanking region through the EC6 and considerable divergence thereafter, as expected from the cDNA sequence (Table 3). Although a microsatellite repeat consisting of the trinucleotide AAG was observed in the intron between the ECmp and TM exons of the *Pirb* gene (14), there were no such repeats in the introns of the *Pira6* gene. However, LINE and SINE repetitive DNA sequences were identified in both *Pirb* and *Pira6* genes (23). A retroviral transposon reverse transcriptase related sequence was found in the 3′ end of *Pira6* gene (∼10,689–11,287). A Blast search of the mouse genome (http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html) identified homologous sequences (> 82% nucleotide identity) that were distributed throughout the genome except for chromosomes 19 and Y.

Overview of the *Pir* locus

There are substantial gaps in the public and Celera sequence databases around the centromeric region of mouse chromosome 7 where the *Pir* genes have been mapped. However, based on our analysis of currently available sequence, the *Pir* genes are distributed in three clusters at the proximal end of chromosome 7 **(**Fig. 4**)**. The first centromeric cluster includes three genes (*Pirb*, *Pira5*, *Pira3*) in the same transcriptional orientation. The next cluster contains only a single *Pira* gene in the opposite orientation, and is almost identical (> 99%) to the 129/OLA PIR-A6 cDNA. Two predicted pseudogenes, designated ψ*Pir1* and ψ*Pir2*, with weak amino acid sequence homology ($E = 4e^{-11}$ and $E = 4e^{-20}$, respectively) to the rat PIR were found to flank both these *Pir* clusters. Two predicted genes with homology to the *Ilt/Lir* were also identified. One gene, *Lirh1*, is predicted to encode a protein homologous ($E = 6e^{-53}$) to a hypothetical mouse protein with homology ($E = 3e^{-21}$) to the chimpanzee LIRe (Accession number AAL31878). *Lirh1* is also homologous to mouse *Pir* ($E = 1e^{-29}$). A second gene, *Lirh2*, encodes a protein most homologous ($E = 1e^{-25}$) to ILT11. The orientation of the *Pir* genes in these two clusters and the location of the flanking genes in the region extending from protein kinase Cγ (*Pkcg*) to at least the interleukin 11 (*IL-11*) gene were found to be remarkably similar to those of *Ilt/Lir* and their flanking genes on human chromosome 19q13.4 (19). However, unlike the human *Ilt/Lir* locus which is composed of two gene clusters, a third *Pir* gene cluster was identified in a region ∼4 mb distal to cluster two. This region contained two additional *Pira* genes that had no exact match with the previously reported BALB/c-derived PIR-A cDNA clones and may be novel *Pir* genes. Moreover, two other incomplete *Pira* sequences corresponding to the 3′ end of the genes were identified in a region distal to the third *Pir* cluster.

Discussion

The variable PIR-A cDNA clones identified in our previous studies were predicted to be encoded by multiple *Pira* genes, but genomic DNA blot analysis using a PIR-A specific probe revealed a relatively simple band pattern, suggesting conservation of restriction enzyme sites flanking the *Pira* genes (1). In the present study, we used P1 clones of 129/OLA mouse genomic DNA in a long range PCR to characterize the exon/intron organization of *Pira* genes. The P16355 P1 clone with an insert DNA of ∼90 kb was found to contain only one *Pira* gene, *Pira6*. Like the *Pirb* gene (14, 15), the *Pira6* gene was found to be relatively small (∼8.2 kb), and ∼58% of the gene comprised intronic sequences. These findings indicate that *Pir* belongs to the type of genes with short introns. The human PIR homolog *Ilt/Lir* genes also have short introns (17, 18). Notably, the average intron to exon length proportion for human genes is ∼26 : 1, with the median being ∼11 : 1 (33, Fyodor A. Kondrashov, personal communication). Given the fact that the transcription process in eukaryotes is slow (∼20 nucleotides transcribed per second) and biologically expensive (at least two ATP molecules per nucleotide), transcription of genes with short introns is less costly than those with long introns, which are particularly common in mammals (34). Interestingly, highly expressed genes have substantially shorter introns than genes expressed at low levels, and this difference is greater in humans than in nematodes (33). In this regard, natural selection appears to favor short introns in highly expressed genes to minimize the cost of transcription and other molecular processes, such as splicing (33). Selection favoring short introns is particularly strong for genes that

Candidate transcription factor binding motifs in the 5′-flanking region of the *Pira6* gene

Transcription factor	St[a]	Sequence (Nucleotide position[b])	Similarity[c]	Cell specificity/function
MZF-1	+	catGGGGa (−105)*[d]	0.97	Important role in the induction of granulopoiesis, linked to the expansion of myeloid precursors
	+	aaaGGGGa (−357)	0.97	prior to terminal differentiation
	+	aatGGGGa (−996)*	0.98	
c-Myb	+	aatatgtactGTTGtgtg (−330)*	0.92	c-Myb expression is crucial for the survival and proliferation of early hemopoietic progenitors,
	+	acctgactctGTTGacag (−704)	0.91	whereas c-Myb down-regulation is required for terminal differentiation.
Ikaros	−	ctgaGGGAttgg (−67)*	0.92	Key role in the decision to differentiate along the lymphoid lineage, disruption of the Ikaros
	+	catgGGGActgc (−105)	0.92	gene leads to severe deficiency in B, T and NK cell development
	+	aaagGGGAacaa (−357)	0.91	
	+	ggaaGGGAaaac (−535)*	0.94	
	−	tcttGGGAtaca (−862)*	0.95	
	+	aagtGGGAtatc (−947)*	0.96	
	+	aatgGGGAggcc (−996)*	0.90	
NF-Y	+	tacCCAAtccc (−70)*	0.93	A ubiquitous heteromeric transcription factor playing an important role in the induction of
	−	tgtCCAAtgtg (−682)	0.91	gene expression in differentiating monocytes
TCF11	+	GTCAtggggactg (−107)*	0.96	TCF11 is widely expressed and vital during embryonic development and has the potential of
	−	GTCAtatatcagc (−123)*	0.98	complex regulation including antioxidant and TNFa responses. Several different isoforms are
	−	GTCAtggccatat (−288)*	0.95	generated by alternative splicing.
	−	GTCAtggggttac (−819)*	0.97	
	+	GTCAtattcgtac (−885)	0.98	
GATA1	+	tgctGATAtatga (−124)*	0.93	GATA1 and GATA2 play key roles in positive regulation of erythroid and megakaryocyte
	+	ggcaGATAtacag (−429)*	0.94	development and negativeregulation of myeloid development. GATA3 expression is
	+	cactGATAtggct (−747)	0.94	restricted to T and NK cells and is required for T cell lineage commitment and differentiation.
	−	cttggGATAcaggt (−865)*	0.95	
	+	ttgaGATAtaaac (−933)*	0.92	
GATA2	+	tgaGATAtaa (−932)	0.94	
AP-4	−	atCAGCagcc (−127)	0.91	AP-4 is ubiquitously expressed and interacts with both promoter and enhancer regions of viral
	−	ggCAGCagac (−722)*	0.91	and cellular origins (e.g., Ig κ promoter E-box).
c-Rel	+	gggcagTTCC (−148)*	0.93	A member of the Rel/NF-kB transcription factors, involved in the late phase of T cell activation.
	−	tggcgtTTCC (−237)*	0.95	Rel deficient mice are defective in mitogenic activation of B and T cells and display impaired
	−	aggatcTTCC (−417)*	0.91	humoral immunity.
AP-1	−	tcTGACttcct (−173)*	0.93	A role in the modulation of apoptotic pathways and the functional development of
(Jun/Fos)	−	tgTGACtaggt (−190)	0.91	monocyte/macrophages, granulocyte, megakaryocyte, mast cells and erythrocytes.
	+	ccTGACctagt (−194)	0.90	
	+	ttTGACacaca (−625)*	0.90	

Table 2

Continued

Transcription factor	St[a]	Sequence (Nucleotide position[b])	Similarity[c]	Cell specificity/function
	+	tcTGACctcca (−645)*	0.90	
	+	ccTGACtctgt (−703)*	0.92	
NF-κB	−	tggcgtTTCC (−237)*	0.92	A central role in regulation of several genes linked with immune, inflammatory and apoptotic processes
NFAT	+	atgagGAAAcgc (−241)*	0.94	Upon T cell activation, NFAT becomes dephosphorylated
	+	gaaggGAAAaca (−534)*	0.95	by calcineurin, translocates into the nucleus and regulates
	+	acaagGAAAact (−674)*	0.97	transcription of various genes (e.g., IL-2, 3, 4, 13, 15, GM-CSF, TNFα, CD40 ligands, Fas ligands).
	+	cttagGAAAatg (−778)*	0.98	
C/EBPβ	−	aacttcaGCAAtca (−663)*	0.93	An important regulatory role in macrophage and B cell development, especially as a regulator of cytokine gene expression during inflammatory responses.

[a]Forward and reverse strands of DNA indicate + and −, respectively
[b]All positions are relative to the ATG translation start site
[c]Values indicate the matrix similarity of transcription binding site, and the core similarities of all sites listed are 1.000
[d*]Indicates that the *Pirb* gene has the same transcription factor binding motifs at similar positions

Table 2

are required to be expressed at high levels at short notice, such as stress-induced proteins. The *Pir* genes, whose products are predicted to be involved in innate immunity, may thus belong to the type of genes that are highly expressed at a relatively short notice.

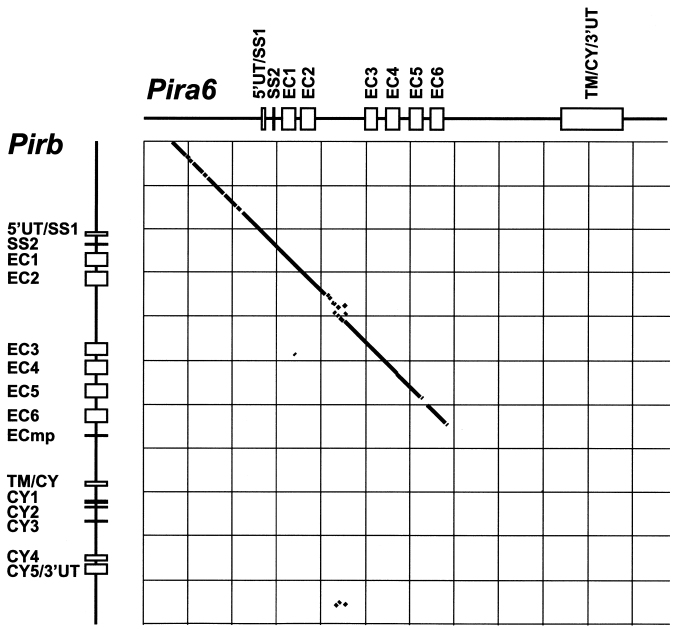

Fig. 3. Dot matrix analysis of *Pira6* and *Pirb* genes. The DNA sequences of *Pira6* (horizontal) and *Pirb* (vertical) were analyzed by the DNA Star Dotplot program using the percentage of 90, the window of 30 and the minimum quality of 10.

Comparison between *Pira6* and *Pirb*[a]

	PIR-A6 Size (bp)	Add	Sub	Del	Differ	%	PIR-B Size (bp)
5′ flank	[2070]	27	101	26	154	93	[2070]
5′UT/SS1	**99**	**0**	**2**	**1**	**3**	**97**	**100**
Int 1	154	0	2	0	2	99	154
SS2	**36**	**0**	**0**	**0**	**0**	**100**	**36**
Int 2	181	0	5	0	5	97	181
EC1	**285**	**0**	**10**	**0**	**10**	**96**	**285**
Int 3	149	0	5	0	5	97	149
EC2	**306**	**0**	**4**	**0**	**4**	**99**	**306**
Int 4	1106	3	50	217	270	76	1320
EC3	**285**	**0**	**7**	**0**	**7**	**98**	**285**
Int 5	126	0	7	0	7	94	126
EC4	**309**	**0**	**21**	**0**	**21**	**93**	**309**
Int 6	228	5	11	0	16	93	223
EC5	**297**	**0**	**5**	**0**	**5**	**98**	**297**
Int 7	235	0	18	25	43	82	260
EC6	**303**	**0**	**6**	**0**	**6**	**98**	**303**
Int 8	[232]	1	58	70	129	44	[301]

[a]Nucleotide sequences of *Pira6* and *Pirb* were aligned by the Clustal method of the DNASTAR MEGALIGN program. Addition (Add), substitution (Sub), deletion (Del) and total differences (Differ) of nucleotides in the indicated regions of *Pira6* by comparing with the *Pirb* sequence are shown. Values in the percentage column indicate the sequence identity (%) with Pirb. For 5′ flanking and intron 8 regions, a part of the nucleotide sequences indicated by numbers in parentheses was used for this comparison

Table 3

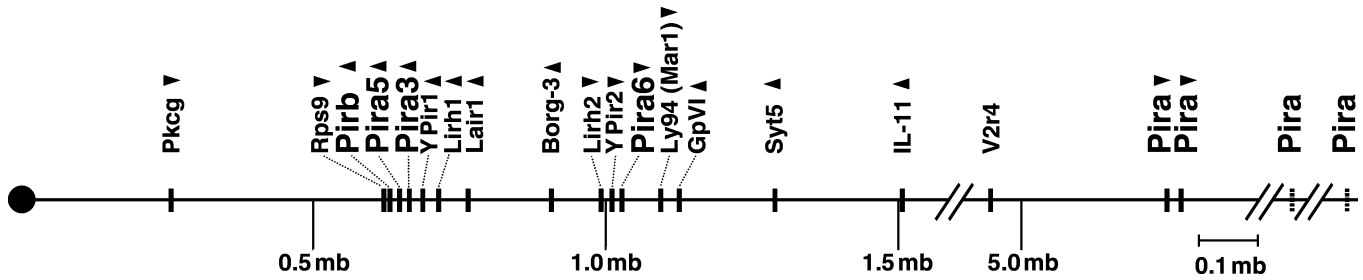

Fig. 4. The *Pir* locus on mouse chromosome 7. Three *Pir* gene clusters, located ~0.6, ~1.0 and ~5.3 mb from the centromere (●) of chromosome 7, and their flanking genes have been identified by analysis of the Celera and NCBI genomic databases and by the cloning of the *Pira6* gene. Gene transcription orientations are indicated by arrowheads (▶ or ◀) except for genes where the transcriptional orientation is unknown. *Pirb* is the first gene in the most centromeric cluster along with *Pira5* and *Pira3*, and *Pira6* appears to exist in isolation, consistent with the finding of only one *Pir* gene on the ~90 kb P16335 clone. The most centromere distal region contains two additional *Pira* genes and two incomplete *Pira* sequences (dashed lines) that were identified in the Celera, but not the NCBI, database. Indicated map distances in megabases (mb) are derived from the Celera database. The flanking genes listed above and the respective NCBI protein accession numbers with the highest homology are: *Pkcg*, protein kinase Cγ (NP_035232); *Rps9*, ribosomal protein s9 (AAH31746); *Lair1*, leukocyte associated immunoglobulin-like receptor 1 (AAM97645); *Borg3*, binder of Rho GTPase3 (NP_067429); *Ly94 (MAR1)*, mouse activating receptor 1 (NP_034876); *gpVI*, platelet glycoprotein VI (XP_145298); *Syt5*, synaptotagmin 5 (NP_058604); *IL11*, interleukin 11 (NP_032376); *V2r4*, vomeronasal organ family 2, receptor 4 (NP_033519).

Based on differences in length of the membrane proximal extracellular region, two types of PIR-A cDNA clones were previously identified: one encoding a longer (28 aa) ECmp and the other encoding a shorter (11 aa) ECmp. We initially considered that this difference might result from alternative splicing of a mini-exon encoding a 17-aa ECmp, but the data from the present genomic analysis indicates that the difference is due to the presence or absence of a functional mini-exon. The PIR-A6 cDNA encodes a shorter ECmp, and the *Pira6* gene lacks a functional mini-exon, because the putative mini-exon is flanked at its 3′ end with 'at', instead of 'gt', and is thereby incapable of serving as the 5′ splice donor signal to the TM/CY/3′UT exon. Interestingly, the change of 'g' to 'a' nucleotide at the 3′ end of the putative mini-exon was also found in other *Pira* genes (four out of seven), indicating that this nucleotide change is not unique for *Pira6*. As this putative mini-exon is flanked at its 5′ end with 'ag' as the 3′ splice acceptor signal, it is possible that alternative splicing may occur between the EC6 exon and this putative exon, resulting in a generation of soluble form of PIR-A6 which consists of six Ig-like domains and a hydrophilic carboxyl tail of 33 aa. This possibility is consistent with our recent finding indicating that soluble PIR molecules of ~70 kDa are present in sera of various strains of mice at concentrations ranging from 10 to 50 ng/mL (Dong-Won Kang and HK, unpublished). A soluble form of ILT/LIR has also been reported (35). It remains unresolved whether these serum PIR proteins are products of the alternatively spliced PIR-A transcripts or result from proteolytic cleavage of cell surface PIR proteins.

Comparison between the *Pira6* and *Pirb* genes revealed extensive homology from ~2 kb 5′ flanking region of the ATG initiation site to the beginning of intron 8 (i.e., downstream of the EC6 exon) and considerable divergence thereafter. This genomic structural feature is also preserved in the genes encoding the activating and inhibitory isoforms of the human PIR homolog ILT/LIR (16–18). Another similarity in the genomic organization of *Pir* and *Ilt/Lir* is that the transmembrane, cytoplasmic and 3′ UT regions of the activating isoforms are encoded by a single exon, whereas the corresponding regions of the inhibitory isoforms are encoded by six exons. One clear distinction between *Pir* and *Ilt/Lir* is that the ECmp regions are encoded by two mini-exons for both inhibitory and activating isoforms of the *Ilt/Lir* genes, but by a single mini-exon for the inhibitory *Pirb* gene and some of the activating *Pira* genes. The remaining *Pira* genes, like *Pira6*, lack the ECmp mini-exon. The number of exons encoding Ig-like domains also differs between *Pir* and *Ilt/Lir*: six for *Pir* and four for *Ilt/Lir*. Intriguingly, two predicted genes (*Lirh1* and *Lirh2*) similar to an *Ilt/Lir* gene in the chimpanzee (Pan troglodytes) and ILT-11 in humans were found during the analysis of Celera and NCBI databases, but whether these genes indeed encode a protein remains unclear.

The alignments of multiple *Ilt/Lir* genes have been established in relation to other members of the LRC (e.g., NKp46, FcαR, KIR and LAIR) on human chromosome 19q13.4 (17–19). The *Ilt/Lir* genes consist of two inverted, duplicated clusters separated by ~200 kb of DNA including *Lair* and other genes. Each of the two clusters spans ~150 kb of genomic sequence immediately centromeric to the *Kir* genes. Together they comprise 13 *Ilt/Lir* genes (17, 19). In contrast, the alignments of multiple *Pir* genes on mouse chromosome 7 have not yet been determined. Based on the present results and the available data in the Celera and NCBI mouse genome database, it appears that the *Pir* genes are located in three regions of the proximal end of chromosome 7. The middle region contains only a single *Pir* gene that is almost identical (> 99%) to the 129/OLA *Pira6* described here,

which may account for the existence of only one *Pir* gene on the P1 clone we analyzed. The most centromeric region includes three *Pir* genes (*Pirb*, *Pira5*, *Pira3*) in an opposite orientation to the other *Pir* genes, similar to the situation with the human *Ilt/Lir* genes (17–19). The flanking genes and their transcriptional orientations are also well conserved between mice and humans. Interestingly, unlike the human *Ilt/Lir* with only two gene clusters, a third *Pir* cluster was identified in the most centromere distal region where two additional *Pira* genes and two incomplete *Pira* sequences were found. This cluster of genes was found in the Celera but not the NCBI mouse genome database. A comparison of this region in the two databases revealed that the Celera sequence is nearly 1 mb longer, which would account for the absence of the third *Pir* gene cluster in the NCBI database. Based on the number of unique *Pira* cDNA sequences that have been isolated, additional *Pira* genes must exist in unsequenced regions of the mouse genome. The identification of activating and inhibitory pairs of chicken Ig-like receptors as distant relatives of the mouse *Pir* and human *Ilt/Lir* genes suggests the existence of a common ancestor, prior to separation of the avian and mammalian lineages (36).

The identification of the transcription start sites of *Pir* genes will provide the framework for future studies on gene regulation. Our analysis has revealed that both *Pira* and *Pirb* genes lack TATA- or CAAT-like promoter sequences. Instead, initiator consensus sequences were found to overlap the transcription start sites, a feature also preserved in the *Ilt/Lir* genes (18). Several hematopoietic transcription factor binding sites that flank both *Pira6* and *Pirb* genes include MZF-1, c-Myb, Ikaros, NF-Y, GATA and C/EBPβ, which are likely candidates for regulating the coordinate expression of these genes (22, 30–32), consistent with their original designation as 'paired' receptors (1). PIR-A and PIR-B receptors are expressed by many hematopoietic cell types, including B lymphocytes, dendritic cells, monocyte/macrophages, granulocytes, mast cells, and megakaryocyte/platelets and the relative surface PIR levels increase as a function of cellular maturation and activation (4). However, it is now evident that in addition to these mature cells, PIR receptors are also expressed by hematopoietic progenitors as defined by the lack of expression of lineage markers and the expression of c-kit and Sca-1 antigens (Ching-Cheng Chen et al unpublished). How *Pir* gene expression is regulated at different stages of hematopoietic cell differentiation remains unclear, but the sequence data of the 5′ flanking regions of *Pirb* (14) and *Pira6* described in the present paper would be the first step in addressing this issue.

References

1. Kubagawa H, Burrows PD, Cooper MD. A novel pair of immunoglobulin-like receptors expressed by B cells and myeloid cells. *Proc Natl Acad Sci U S A* 1997: **94**: 5261–6.
2. Hayami K, Fukuta D, Nishikawa Y et al. Molecular cloning of a novel murine cell-surface glycoprotein homologous to killer cell inhibitory receptors. *J Biol Chem* 1997: **272**: 7320–7.
3. Maeda A, Kurosaki M, Kurosaki T. Paired immunoglobulin-like receptor (PIR)-A is involved in activating mast cells through its association with Fc receptor γ chain. *J Exp Med* 1998: **188**: 991–5.
4. Kubagawa H, Chen CC, Ho LL et al. Biochemical nature and cellular distribution of the paired immunoglobulin-like receptors, PIR-A and PIR-B. *J Exp Med* 1999: **189**: 309–17.
5. Taylor LS, McVicar DW. Functional association of FcεRIγ with arginine^{632} of paired immunoglobulin-like receptor (PIR)-A3 in murine macrophages. *Blood* 1999: **94**: 1790–6.
6. Ono M, Yuasa T, Ra C, Takai T. Stimulatory function of paired immunoglobulin-like receptor-A in mast cell line by associating with subunits common to Fc receptors. *J Biol Chem* 1999: **274**: 30288–96.
7. Bléry M, Kubagawa H, Chen CC, Vély F, Cooper MD, Vivier E. The paired Ig-like receptor PIR-B is an inhibitory receptor that recruits the protein-tyrosine phosphatase SHP-1. *Proc Natl Acad Sci U S A* 1998: **95**: 2446–51.
8. Maeda A, Kurosaki M, Ono M, Takai T, Kurosaki T. Requirement of SH2-containing protein tyrosine phosphatases SHP-1 and SHP-2 for paired immunoglobulin-like receptor B (PIR-B) -mediated inhibitory signal. *J Exp Med* 1998: **187**: 1355–60.
9. Yamashita Y, Ono M, Takai T. Inhibitory and stimulatory function of paired Ig-like receptor (PIR) family in RBL-2H3 cells. *J Immunol* 1998: **161**: 4042–7.
10. Uehara T, Bléry M, Kang DW et al. Inhibition of IgE-mediated mast cell activation by the paired Ig-like receptor PIR-B. *J Clin Invest* 2001: **108**: 1041–50.
11. Kubagawa H, Cooper MD, Chen CC et al. Paired immunoglobulin-like receptors of activating and inhibitory types. *Curr Top Microbiol Immunol* 1999: **244**: 137–49.
12. Takai T, Ono M. Activating and inhibitory nature of the murine paired immunoglobulin-like receptor family. *Immunol Rev* 2001: **181**: 215–22.
13. Ujike A, Takeda K, Nakamura A, Ebihara S, Akiyama K, Takai T. Impaired dendritic cell maturation and increased TH2 responses in PIR-B$^{-/-}$ mice. *Nature Immunol* 2002: **3**: 542–8.
14. Alley TL, Cooper MD, Chen M, Kubagawa H. Genomic structure of PIR-B, the inhibitory member of the paired immunoglobulin-like receptor gene in mice. *Tissue Antigens* 1998: **51**: 224–31.
15. Yamashita Y, Fukuta D, Tsuji A et al. Genomic structures and chromosomal location of p91, a novel murine regulatory receptor family. *J Biochem* 1998: **123**: 358–68.
16. Wagtmann N, Rojo S, Eichler E, Mohrenweiser H, Long EO. A new human gene complex encoding the killer cell inhibitory receptors and related monocyte/macrophage receptors. *Curr Biol* 1997: **7**: 615–8.
17. Wilson MJ, Torkar M, Haude A et al. Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc Natl Acad Sci U S A* 2000: **97**: 4778–83.

18. Liu WR, Kim J, Nwankwo C, Ashworth LK, Arm JP. Genomic organization of the human leukocyte immunoglobulin-like receptors within the leukocyte receptor complex on chromosome 19q13.4. *Immunogent* 2000: **51**: 659–69.
19. Barten R, Torkar M, Haude A, Trowsdale J, Wilson MJ. Divergent and convergent evolution of NK-cell receptors. *Trends Immunol* 2001: **22**: 52–7.
20. Martin AM, Kulski JK, Witt C, Pontarotti P, Christiansen FT. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol* 2002: **23**: 81–8.
21. Sternberg N, Smoller D, Braden T. Three new developments in P1 cloning: increased cloning efficiency, improved clone recovery, and a new P1 mouse library. *Genet Anal Techn* 1994: **11**: 171–80.
22. Quandt K, Frech K, Karas H, Wingender E, Werner T. MatInd and MatInspector – new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucl Acids Res* 1995: **23**: 4878–84.
23. Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR – a program for identification and elimination of repetitive elements from DNA sequences. *Comp Chem* 1996: **20**: 119–22.
24. Shaw G, Kamen R. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* 1986: **46**: 659–67.
25. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982: **157**: 105–32.
26. Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN. Prediction of potential GPI-modification sites in proprotein sequences. *Protein Engineering* 1999: **12**: 387–94.
27. Eisenhaber B, Bork P, Eisenhaber F. Prediction of potential GPI-modification sites in proprotein sequences. *J Mol Biol* 1999: **292**: 741–58.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990: **215**: 403–10.
29. Ernst P, Smale ST. Combinatorial regulation of transcription I. General aspects of transcriptional control. *Immunity* 1995: **2**: 311–9.
30. Barreda DR, Belosevic M. Transcriptional regulation of hemopoiesis. *Dev Comp Immunol* 2001: **25**: 763–89.
31. Nagamura-Inoue T, Tamura T, Ozato K. Transcription factors that regulate growth and differentiation of myeloid cells. *Intern Rev Immunol* 2001: **20**: 83–105.
32. Faisst S, Meyer S. Compilation of vertebrate-encoded transcription factors. *Nucl Acids Res* 1992: **20**: 3–26.
33. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA. Selection for short introns in highly expressed genes. *Nature Genet* 2002: **31**: 415–8.
34. Ucker DS, Yamamoto KR. Early events in the stimulation of mammary tumor virus RNA synthesis by glucocorticoids. *J Biol Chem* 1984: **259**: 7416–20.
35. Borges L, Hsu ML, Fanger N, Kubin M, Cosman D. A family of human lymphoid and myeloid Ig-like receptors, some of which bind to MHC class I molecules. *J Immunol* 1997: **159**: 5192–6.
36. Dennis G, Kubagawa H, Cooper MD. Paired Ig-like receptor homologs in birds and mammals share a common ancestor with mammalian Fc receptors. *Proc Natl Acad Sci U S A* 2000: **97**: 13245–50.